

Docket Code: AP.PRE.REQ

PTO/SB/33 (07-05)

Approved for use through xx/xx/200x. OMB 0651-00xx

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

## PRE-APPEAL BRIEF REQUEST FOR REVIEW

Docket Number (Optional)

iSPHERES 1

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to "Mail Stop AF, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450" [37 CFR 1.8(a)]

on January 9, 2008

Signature Theresa L. Belich

Typed or printed name Theresa L. Belich

Application Number

09/728,689

Filed

Dec. 1, 2000

First Named Inventor

Eric T. Bax

Art Unit

2178

Examiner

G. Vaughn

Applicant requests review of the final rejection in the above-identified application. No amendments are being filed with this request.

This request is being filed with a notice of appeal.

The review is requested for the reason(s) stated on the attached sheet(s).  
Note: No more than five (5) pages may be provided.

I am the

- ☐ applicant/inventor.
- ☐ assignee of record of the entire interest.  
See 37 CFR 3.71. Statement under 37 CFR 3.73(b) is enclosed.  
(Form PTO/SB/96)

☒ attorney or agent of record.  
Registration number 29355

☐ attorney or agent acting under 37 CFR 1.34.  
Registration number if acting under 37 CFR 1.34 \_\_\_\_\_

David Volejnick  
Signature

DAVID VOLEJNICK  
Typed or printed name

303-938-4154  
Telephone number

1-9-08  
Date

NOTE: Signatures of all the inventors or assignees of record of the entire interest or their representative(s) are required.  
Submit multiple forms if more than one signature is required, see below\*.

☒ \*Total of 3 forms are submitted. Notice, Request (SB/33), Pre-Appeal Brief Request

This collection of information is required by 35 U.S.C. 132. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11, 1.14 and 41.6. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Mail Stop AF, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

IN THE UNITED STATES  
PATENT AND TRADEMARK OFFICE



**Patent Application**

Inventors: **Eric T. Bax**  
**Charles C. Fowlkes**  
**Louis Cisnero, Jr.**

Case No.: **iSpheres 1**

Serial No.: **09/728,689**      Group Art Unit: **2178**

Filing Date: **December 1, 2000**

Examiner: **Gregory J. Vaughn**

Title: **Technique for Extracting Data from Structured Document**

Commissioner For Patents  
PO Box 1450  
Alexandria, VA 22313-1450

**PRE-APPEAL BRIEF REQUEST FOR REVIEW**

SIR:

Claims 43-46 are pending. Claims 43-46 are rejected under 35 U.S.C. §102(e) over U.S. patent no. 6,424,980 (Iizuka et al.). This rejection is respectfully traversed.

This application is directed to extracting of data records from structured text, such as a web page or any text-containing file, without prior knowledge of the structure of the text. The invention deduces the structure of the text by using information about the attributes and knowledge of candidate structures. (Specification at page 3, lines 9-15.) The claims recite how this is accomplished; in other words, what is claimed is not that data records are extracted without prior knowledge of the structure, but how that is effected.

Iizuka et al. represent the prior art referenced by applicants in their "Description of Related Art". On page 1, line 20, to page 2, line 7, applicants state:

Traditional techniques for solving this information gathering problem are typically based on knowledge of the structure used to arrange data within each specific website. (The structure used to arrange the data within a page is commonly referred to as the syntax of the page.) These techniques require prior determination of the syntax of each page and storage of syntax information about each page in a data storage device, such as a database.

When gathering information about a subject from a particular page, the traditional techniques identify the attributes of the subject by comparing the structure of the page with the stored structure information. When there is a match, the traditional technique returns the attribute value to the user.

These traditional techniques are limited because they can only gather attribute values from a page when they know the syntax of a page. To put it differently, the traditional techniques can only gather attribute values when the syntax of a page has been previously determined and stored.

Correspondingly, Iizuka et al. state:

The apparatus has a HTML document storing unit for storing meta data about HTML documents. That meta data includes the locations, document structures, presentation locations, presentation styles, etc., of the HTML documents for each HTML document.... The document structure data of the HTML documents specifies the structures of partial structure such as tables, lists and clauses contained in the HTML documents and is used to map element data in the table and lists to items to be extracted.

(Col. 11, line 63, to col. 12, line 5)

In the preparatory phase, a managing person prepares meta data about HTML documents through the HTML document meta data manager before starting the execution phase.

(Col. 14, lines 30-32).

In other words, Iizuka et al. require the syntax of documents that are to be searched to be known and stored before a search can be conducted.

The Examiner asserted that this argument is not persuasive because applicants' claims do not recite that data records within a file are identified without using prior knowledge of the structure (e.g., syntax) of the file. This assertion misses the point of applicants' argument. Applicants' argument explains why a teaching of how data records within a document may be identified without knowledge of the structure of a file is not found in Iizuka et al.: since Iizuka et al. know the structure a-priori from pre-stored meta-data (see, Iizuka et al. Figs. 12 and 13, and col. 14, lines 17-21), they do not need to identify it, and consequently they do not teach how it may be identified. In contrast, and as was pointed out above, applicants' claims recite how this identification (and consequent record extraction) is done. Applicants are relying on the functionality – the particular steps that are recited in the claims – to distinguish their invention from Iizuka et al. Iizuka et al. do not disclose, teach, or suggest that functionality.

Inter alia, applicants' claims recite "identifying potential locations of values of record fields in [a structured] text by identifying locations in the text of items in lists of known potential values for record fields." The Examiner asserted that "Iizuka discloses identifying potential locations of values of record fields in the text in Figure 8 at reference signs S200." The Examiner is mistaken. This step of Fig. 8 refers to determining the addresses of documents that are to be searched, in an HTML document table that stores the locations of HTML documents -- see col. 14, lines 15-17 and 47-51. In contrast, the claim language refers to identifying locations

of known potential values for record fields, within a document (text) that is being searched.

The Examiner dismissed this argument by stating that “the examiner equates ‘list of known potential values for record fields’ with ‘HTML document table.’” But applicants’ argument cannot be dismissed so easily. The issue is what is being identified. lizuka et al. identify addresses of documents that are to be searched, whereas the claims search a text to identify therein potential values (i.e., “items in lists of known potential values”) for record fields. The things that are being identified in lizuka et al. and in applicants’ claims are unmistakably different.

Applicants’ claims further recite “identifying a region of interest in the text by applying multiple candidate region partitioners, evaluating each to measure how well it isolates a region with a high density and a high amount of potential locations of values of record fields, selecting one that measures best, and applying it to produce a region of interest.” The Examiner asserted that “lizuka discloses identifying a region of interest in the text by applying candidate region partitions and segmenting the region of interest into record regions that contain data for a single record,” and pointed to lizuka et al.’s description of Ashish and Knoblock’s technique at col. 2, lines 45-65, as supporting this assertion. The Examiner is again mistaken. This technique identifies the regions (the internal structure) of a text (document). But it does not identify a region of interest among the regions of the text, as required by applicants’ claims.

Undaunted, the Examiner dismissed this argument by asserting that “lizuka discloses ‘This technique considers a portion in HTML document as meaningful information’ (column 2, line 50, emphasis added). The examiner equates ‘a region of interest’ with ‘portion in HTML document as meaningful information.’” The Examiner’s assertion misses the mark. The statement in lizuka et al. that “this technique considers a portion in HTML document as meaningful information” merely means that portions of an


HTML document are not meaningless -- in other words, that portions, and not only the document as a whole, have meaning. Thus, this statement provides a rationale for why someone would want to determine what portions (regions) a text has. But, significantly, it does not teach identifying a region of interest among those regions, as required by the claims. Nor does it teach identifying the region by the applying, evaluating, and selecting that are recited in the claims. Nor does it teach segmenting the region of interest (as opposed to the text as a whole), as required by the claims. Nor does it effect segmentation by the applying, evaluating, selecting, applying, and extracting that are recited in the claims.

It should therefore be evident that, contrary to the Examiner's assertion, Iizuka et al. do not disclose, teach, or suggest identifying a region of interest in the text as that identifying is recited in the claims. Nor do they disclose the recited segmenting

For the reasons stated above, Applicants request that the Section 102(e) rejection of their claims over Iizuka et al. be reversed.

Respectfully submitted,

**Eric T. Bax**  
**Charles C. Fowlkes**  
**Louis Cisnero, Jr.**

By   
David Volejnicek  
Corporate Counsel  
Reg. No. 29355  
303-538-4154

Date: 1-9-08  
Avaya Inc.  
Docket Administrator  
307 Middletown-Lincroft Road  
Room 1N-391  
Lincroft, NJ 07738